**Minerva Foods**

**Ambiente e acessos**

**Documento de Implementação Técnica**

**DIT-2024-00.01**

**Sumário**

[Histórico de Revisões 3](#_Toc175154232)

[Objetivo do documento 4](#_Toc175154233)

[Acessos necessários – Ambiente On-Premisse 4](#_Toc175154234)

[1. VPN 4](#_Toc175154235)

[2. Apache Guacamole 5](#_Toc175154236)

[3. Air flow 6](#_Toc175154237)

[4. Pentaho Data Integration – PDI 6](#_Toc175154238)

[5. Bitbucket 7](#_Toc175154239)

[6. Banco de dados – SQL 10](#_Toc175154240)

[7. E-mail 11](#_Toc175154241)

[8. Microsoft Azure Synapse 11](#_Toc175154242)

[Arquitetura Implementada 12](#_Toc175154243)

[1. On – Premisse 12](#_Toc175154244)

[ Origem 12](#_Toc175154245)

[ Dw\_trusted 12](#_Toc175154246)

[ Dw\_refined 13](#_Toc175154247)

[ Visualização 13](#_Toc175154248)

[2. Data Lake 13](#_Toc175154249)

# Histórico de Revisões

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Versão** | **Descrição** | **Autor** |
| 18/07/2024 | 0 | Criação do Documento | Jhonatan Benevenuto |
| 21/08/2024 | 1 | Atualização do documento | Jhonatan Benevenuto |

# Objetivo do documento

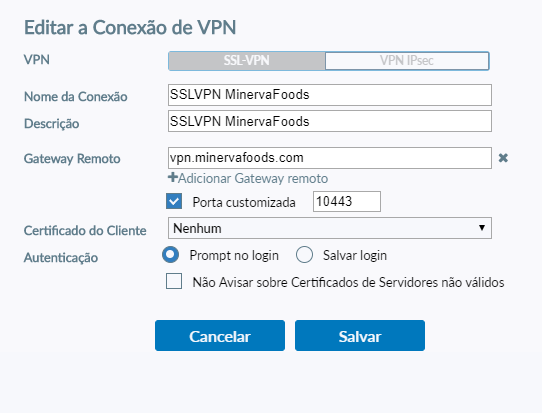
Este documento tem como finalidade documentar quais são os acessos necessários para o time de back-end do cliente Minerva Foods. Tendo o objetivo de facilitar o processo de solicitação de acesso ao ambiente do cliente e as respectivas ferramentas necessárias para que o engenheiro de dados responsável, possa desenvolver soluções funcionais.

# Acessos necessários – Ambiente On-Premisse

## VPN

Esse é o primeiro acesso necessário para conectarmos no ambiente da Minerva, através dele poderemos fazer a conexão nas demais ferramentas que se fazem necessárias para o desenvolvimento do time de engenharia.

O acesso é realizado através da FortiClient VPN com as seguintes configurações:



Se faz necessário Login e senha e a instalação do app mobile FortiToken. Todas essas credenciais são disponibilizadas pelo time da MinervaFoods.

Importante lembrar que essas configurações possuem validade, e se não configurada dentro do prazo, expira.

## Apache Guacamole

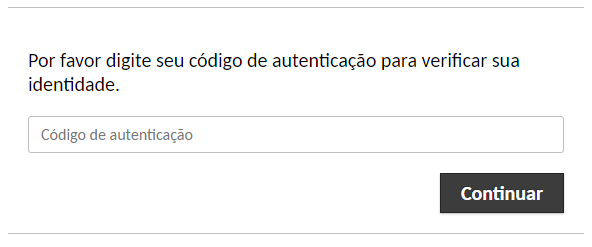
O Apache Guacamole é um gateway de área de trabalho remota sem cliente que suporta protocolos padrão como VNC, RDP e SSH. Como não tem cliente, seus usuários não precisam instalar nada. Eles apenas usam um navegador da Web para acessar remotamente suas VMs.

Através desse acesso é possível se conectar em todo o ambiente da MinervaFoods, pois a partir dele o usuário está dentro da rede do cliente, como se estivesse localmente no cliente. Esse acesso é feito via navegador de internet (Chrome, Edge, etc) através do seguinte link de acesso: <https://guacamole.minervafoods.com:8443/guacamole/#/> .



Para se conectar é necessário estar conectado na VPN, onde o usuário e senha de acesso são os mesmos.

Após se conectar, é solicitado um token de autenticação que é fornecido através do aplicativo Google Authenticator, a configuração desse token, é feita através de um QrCode que o cliente nos manda.



## Air flow

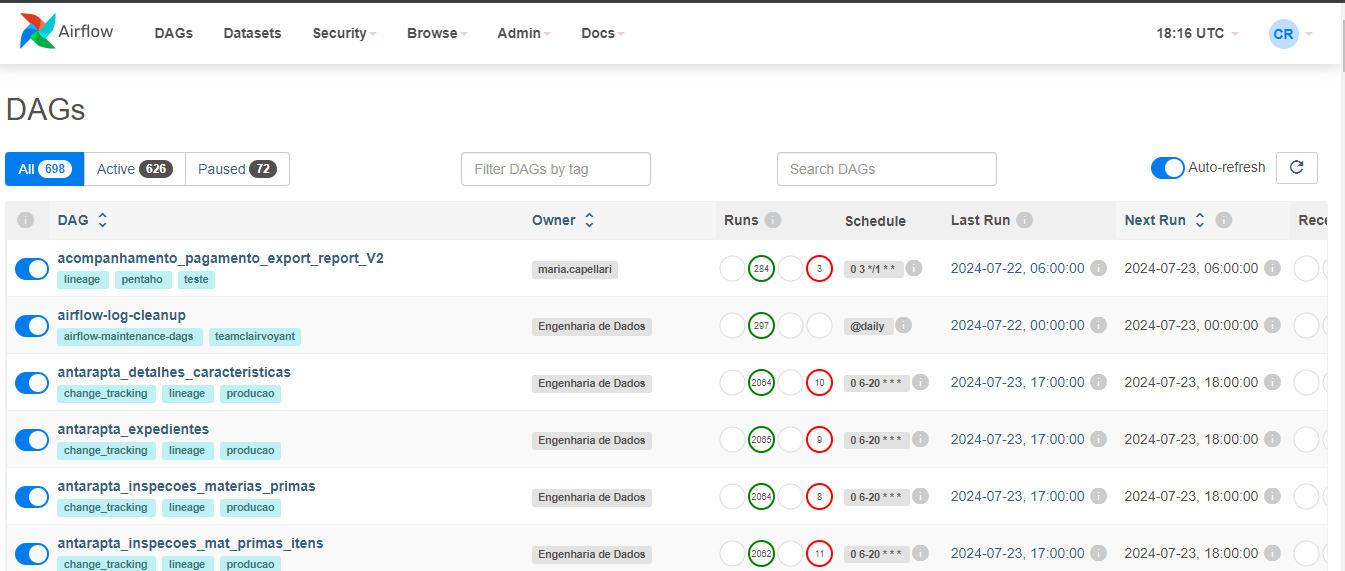
A ferramenta Air flow é utilizada com o Schedule principal da empresa, quando falamos do ambiente on-premisse. É através dele que os pipelines desenvolvidos em pentaho e em Python são executados.

O acesso a ferramenta é através do link:

<http://172.25.2.107:8100/home>

Para se conectar, é necessário que esteja conectado na VPN ,e o acesso é feito pelo browser da máquina pessoal ou através do Guacamole.

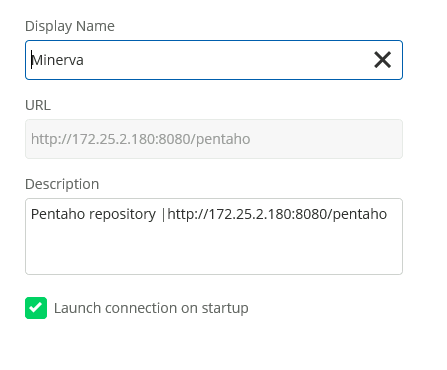
O acesso é liberado pelo cliente, e o usuário e senha são os mesmos da VPN.



## Pentaho Data Integration – PDI

A ferramenta é utilizada para fazer ingestões e transformações no ambiente do cliente. Vários projetos são orquestrados através dessa ferramenta, como por exemplo: IC Comercial.

Para utilizar essa ferramenta é necessário estar conectado na VPN, e utilizar a versão mínima 9.1 do Pentaho Data Integration. Para se conectar, é preciso configurar o repositório com as seguintes informações:



Após essa configuração será solicitado um usuário e senha, que será informado pelo cliente.

O acesso ao Pentaho, pode ser feito localmente na máquina local, ou através do Guacamole. Sugiro utilizar pelo Guacamole, que apresenta menos erros, e já possui o Pentaho instalado, sendo necessário apenas a configuração do repositório.

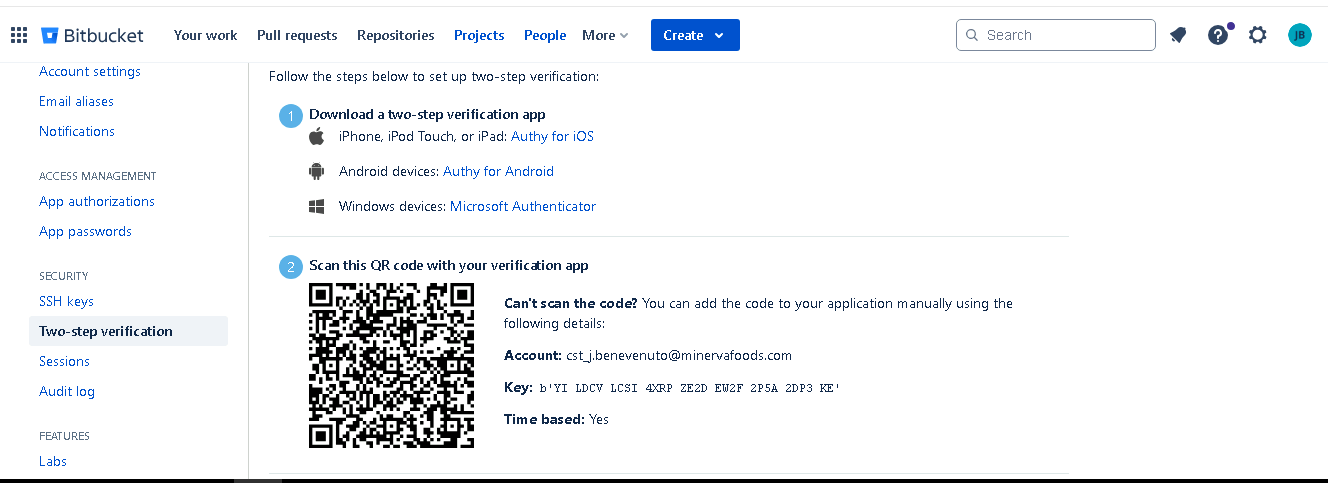
O PDI está localizado na unidade C da VM do Guacamole: C:\data-integration.

## Bitbucket

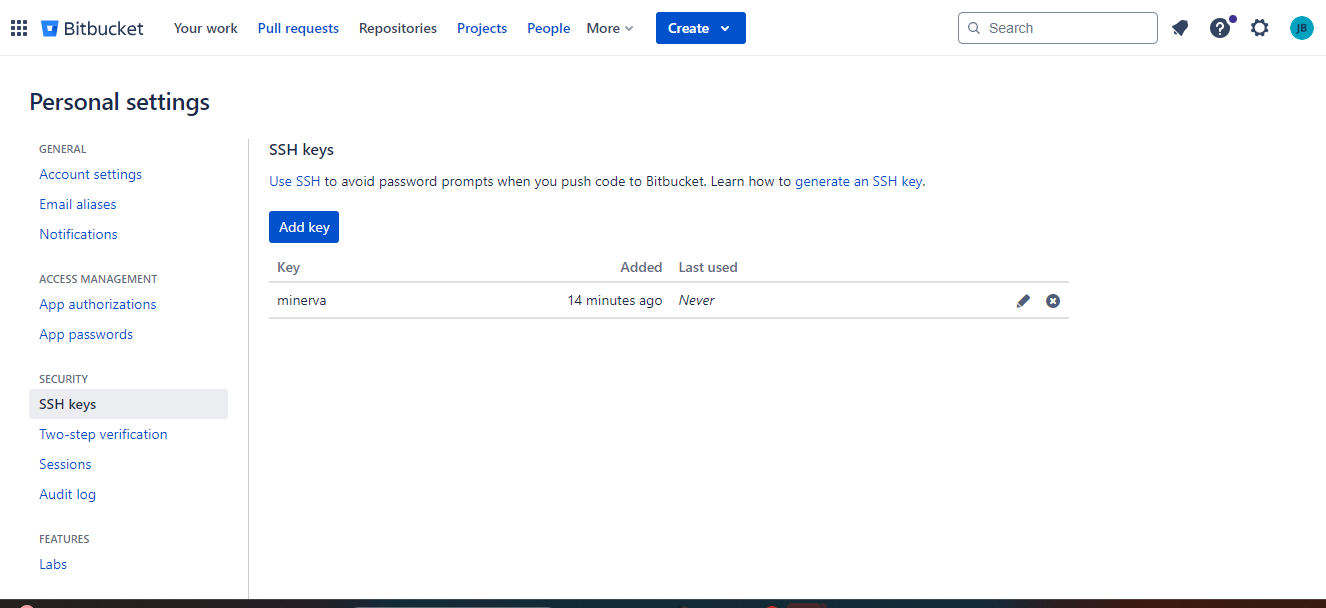
O Bitbucket Cloud é uma ferramenta de hospedagem de código e colaboração baseada em Git, criada para equipes.

O acesso a ferramenta é dado pelo cliente, utilizando o e-mail com o domínio @minervafoods.com.br, através dele o cliente dará permissão ao workspace [minervafoodsdev](https://bitbucket.org/minervafoodsdev/) e aos repositórios minerva\_de\_dags (<https://bitbucket.org/minervafoodsdev/minerva_de_dags/src/master/>) e aos repositórios de desenvolvimento.

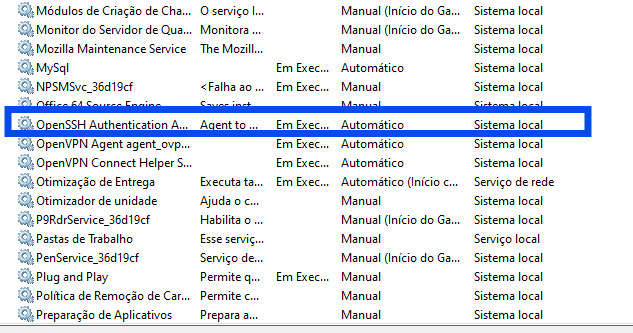
Ao fazer o primeiro acesso será necessário configurar a autenticação de dois fatores, que é feita através do aplicativo Google Authenticator.



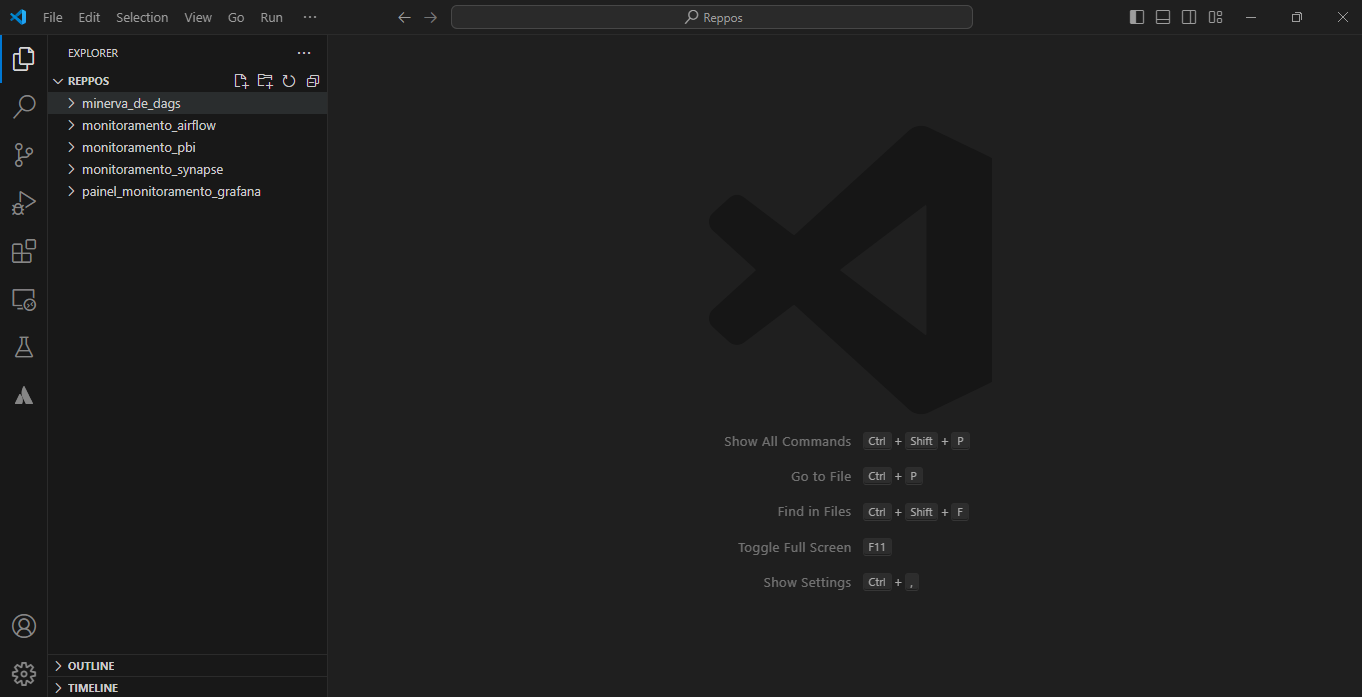
Após acessar, é necessário fazer a clonagem dos repositórios na sua máquina através do VsCode, para fazer o “clone”, é preciso configurar uma chave SSH Keys no menu Personal settings, seguindo as instruções que o próprio Bittbucket apresenta.



Uma dica importante é ter o OpenSSH ativo na sua maquina local, para o funcionamento da clonagem através do SSH.

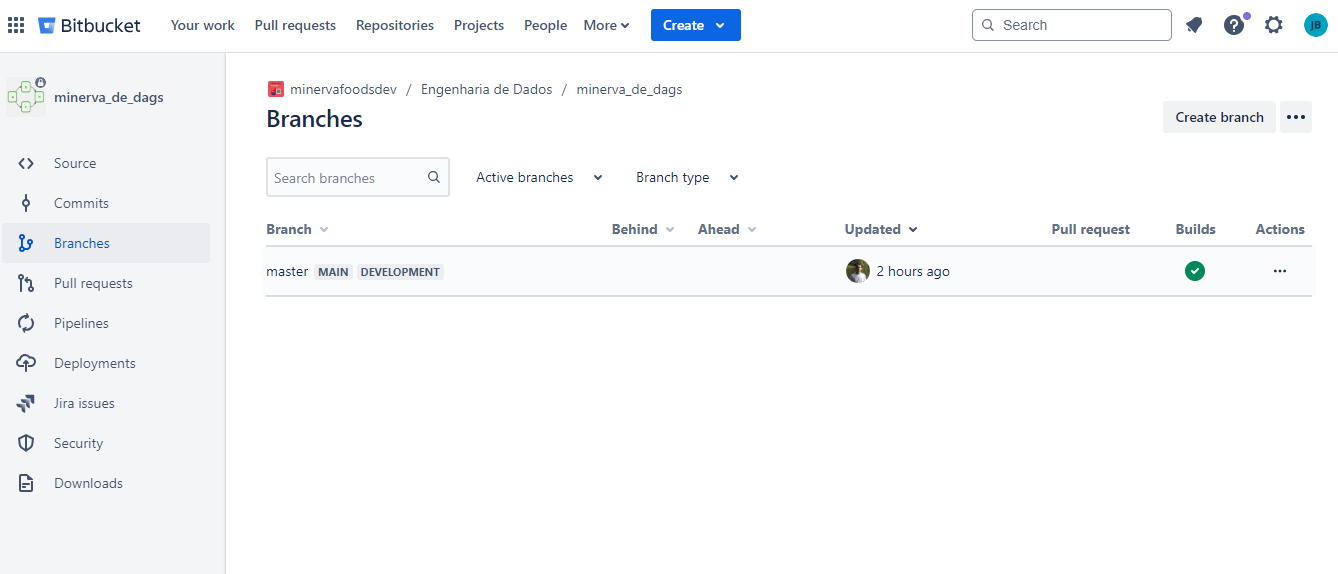


Após essas configurações é só executar a clonagem, apontando para uma pasta que será o seu repositório local dos arquivos.



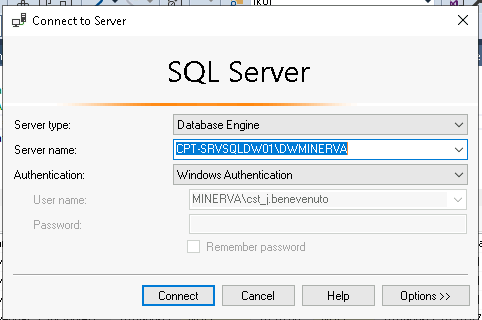
Todo o processo de subida par ao ambiente de produção e versionamento dos códigos é feito pelo BitBucket, seguindo essas etapas:

* Nomenclatura pode seguir algo como cst\_jhonatan\_[atividade].
* Após publicar a branch / fazer as alterações necessárias / commits / o desenvolvedor deve comunicar com o Fred, através do email [frederico.cotrim@minervafoods.com](mailto:frederico.cotrim@minervafoods.com) para aprovação e merge na máster.



## Banco de dados – SQL

Para acessar o Banco de Dados da CPT-SRVSQLDW01\DWMINERVA, é necessário a liberação de acessos do time Minerva. Após a liberação, o acesso é realizado através do GUACAMOLE, acessando a maquina Virtual o usuário para acesso ao banco é feito com o MESMO usuário e senha da VPN.



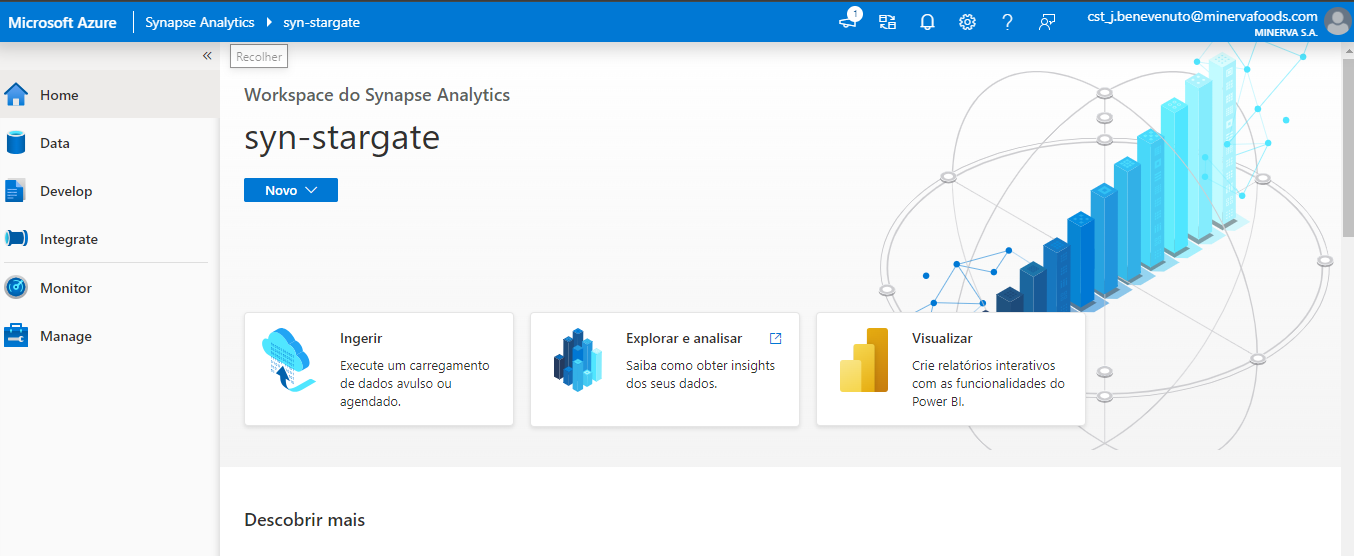
## E-mail

O E-mail @minervafoods.com.br, é configurado pelo time Minerva. O e-mail é utilizado para termos acesso a algumas ferramentas, mas não necessariamente fazemos uso do e-mail, para envio ou recebimento das informações.

## Microsoft Azure Synapse

Azure Synapse Analytics é um serviço de análise empresarial que acelera o tempo de insight entre data warehouses e sistemas de Big Data. Ele reúne o melhor das tecnologias SQL usadas em data warehouse empresarial, tecnologias de Apache Spark para Big Data e Azure Data Explorer análise de log e série temporal.

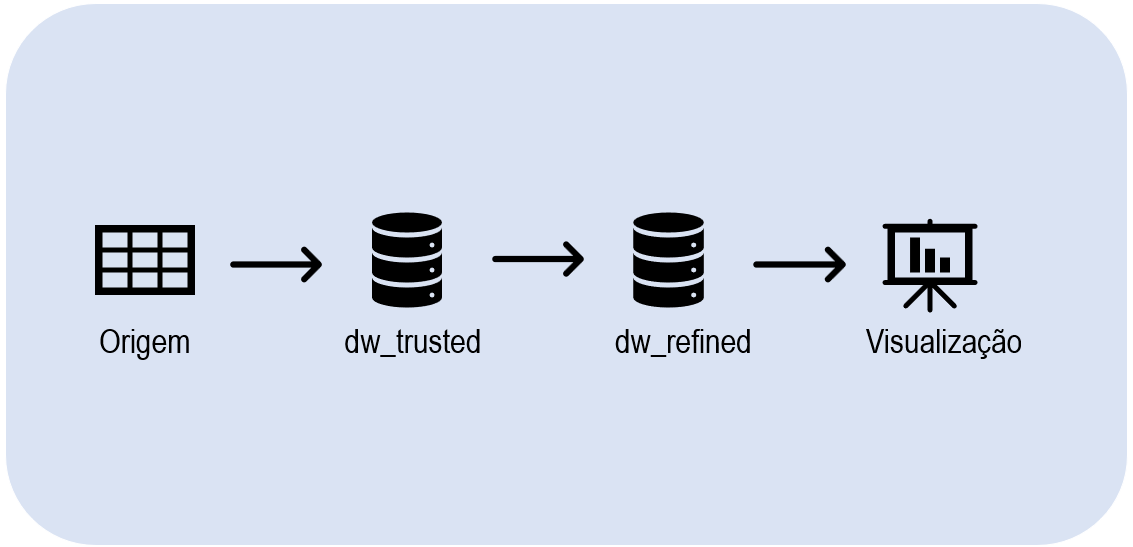
No atual momento não fazemos desenvolvimento nessa ferramenta, mas é importante termos o acesso para podermos confrontar e fazer análise dos pipelines que falham.  
 O acesso é realizado através do e-mail Minerva.



# Arquitetura Implementada

O cliente Minerva Foods possui dois ambientes de dados:

### On – Premisse



### Origem

A Origem dos dados é apontada pelo cliente, podendo ser bancos de dados ou talvez arquivos XLSX, que ficam armazenados no sharepoint do cliente. Um detalhe importante é que para ter acesso a arquivos sharepoint é necessário que o cliente autorize esse acesso, através do usuário de VPN do desenvolvedor.

### Dw\_trusted

O database dw\_trusted fica armazenado no servidor CPT-SRVSQLDW01\DWMINERVA.

O acesso a esse servidor é feito pelo cliente através do usuário de VPN cadastrado e autorizado.

O funcionamento desse database é como se fosse uma camada bronze/prata, é nele que o dado é gravado e recebe algumas modificações necessárias para armazenamento.

O cliente utiliza essa camada como se fosse uma área de staging, só em alguns casos específicos e de extrema necessidade que é utilizado o dw\_staging como camada bronze. Conforme definição do cliente a camada a ser utilizada como staging/bronze é a dw\_trusted.

### Dw\_refined

O database dw\_refined fica armazenado no servidor CPT-SRVSQLDW01\DWMINERVA.

O acesso a esse servidor é feito pelo cliente através do usuário de VPN cadastrado e autorizado

Esse database, é considerado a camada ouro do cliente, nesse local fica armazenado o dado já refinado, e modelado no formato star schema, onde todas as tabelas se relacionam através de surrogate key, seguindo as boas práticas de BI.   
 As tabelas disponibilizadas, devem seguir um padrão de nomenclatura e dicionário de dados, que é definida pelo cliente.

Essa camada é onde os desenvolvedores de data viz, vão fazer o consumo dos dados.

### Visualização

Todos os dados são consumidos em PowerBI, seguindo as boas práticas de desenvolvimento.

### Data Lake

Esse ambiente não tivemos contato ainda, e desconheço.